

# Joint Radio Resource Management for LTE-UMTS Coexistence Scenarios

Nemanja Vučević, Jordi Pérez-Romero, Oriol Sallent, Ramon Agustí

Dept. TSC, Universitat Politècnica de Catalunya (UPC)

c/ Jordi Girona 1-3, 08034, Barcelona, Spain

{vucevic, jorperez, sallent, ramon}@tsc.upc.edu

**Abstract**— Constantly increasing demand for throughput and quality in wireless communication systems leads to continuous research of wise radio resource management, because of the scarce availability of frequency bands and the consequent capacity limitations. In addition, technology evolution is addressed towards spectral efficient techniques that can offer higher data rates. This is the case of OFDMA (Orthogonal Frequency-Division Multiple Access), introduced by 3GPP as the technology for future Long Term Evolution (LTE). However, given the current penetration of legacy technologies such as UMTS (Universal Mobile Telecommunications System), operators will have to deal with the coexistence of multiple Radio Access Technologies (RATs), so that the exploitation of the complementarities between technologies through Joint Radio Resource Management (JRRM) mechanisms will be needed. In this paper we propose a novel dynamic JRRM algorithm for LTE-UMTS coexistence scenarios. The proposed mechanism is based on Reinforcement Learning (RL) which is considered to be a good candidate to achieve cognition in future reconfigurable networks. The proposed solution implements autonomous RL agents in each base station which decide on the allocation of the most suitable RAT to each user. We give a detailed description of the solution and analyze the behavior under various load conditions. We also demonstrate the capability of the algorithm to adjust in dynamic scenarios.

**Keywords:** joint radio resource management, LTE, WCDMA, reinforcement learning

## I. INTRODUCTION

Digital mobile telecommunications penetrated mass markets with voice as a primary service at first. However, development of mobile communications in the last decade is characterized by constant expansion of data services. Lower prices of mobile services brought heterogeneity in every sense: presence of vast variety of mobile services, heterogeneity in Radio Access Technologies (RATs), growth in number of wireless service providers and users' profiles, etc. To achieve capacity maximization in such complex and competitive circumstances, optimized use of spectral resources is needed. Spectral efficiency improvements require, on the one hand, to have evolved technologies that can offer higher data rates and, on the other hand, to operate the network with smart Radio Resource Management (RRM) mechanisms that enable the dynamic use of radio resources to meet the instantaneous users' demands and cope with network dynamics in terms of mobility,

traffic generation, etc. Furthermore, the coexistence of multiple access technologies, due to e.g. the deployment of the different generations of mobile communications, has provided an additional degree of freedom in the RRM through smart mechanisms that take jointly into account the resources available in all the RATs to make the appropriate allocations. These are referred to as Joint RRM (JRRM) or Common RRM (CRRM) [1].

Starting from Release 8, 3GPP has adopted OFDMA (Orthogonal Frequency-Division Multiple Access) as the technology for the so-called Long Term Evolution (LTE) [2] to achieve high traffic capacities in 4G mobile networks. However, the penetration of WCDMA (Wideband Code Division Multiple Access) as 3G UMTS (Universal Mobile Telecommunications System) in mobile markets is already significant. Therefore, it is expected that these two RATs will have to coexist during significant periods of time. During these periods, WCDMA can be used not only to serve those mobile devices that have no support of OFDMA, but also, thanks to the relatively high bit rates and good performance that can be obtained for some services with WCDMA, this technology can be exploited to achieve an efficient use of the total spectrum allocated to UMTS and LTE by a smart distribution of the users that support both technologies.

Under the above consideration, in this paper we present a JRRM solution for scenarios in which LTE and UMTS technologies coexist. The objective of the proposed mechanism is to decide dynamically the most suitable RAT to serve each session. Adaptability in this solution is achieved by applying Reinforcement Learning (RL) actor-critic methods [3]. We choose reinforcement learning as a mechanism to cope with the dynamic changes in the radio environment, providing the network with the necessary cognition capabilities to modify its operation in accordance with the network status [4]. In our previous work, we have applied similar learning technique to improve spectrum efficiency by means of dynamic spectrum management in WCDMA [5].

Different works have addressed the JRRM problem in the literature during the recent years. The JRRM algorithms developed so far apply on different network aspects (see [6] and references within) and usually consider GPRS (General Packet Radio Service), UMTS and WLAN (Wireless Local Area Network) [7]. JRRM solutions for these networks range

---

This work has been supported by the Spanish Research Council and FEDER funds under COGNOS grant (ref. TEC 2007-60985); and MEC-FPU Scholarship (AP2005-3739).

from dynamic load distribution mechanisms [8] to complex Fuzzy-neural [9] solutions for advanced network supervision.

To the authors' best knowledge there is no previous work in literature that addresses JRRM problem for LTE-UMTS coexistence. As for the utilization of machine learning mechanisms, in [10] authors introduce Q-learning model to achieve joint call admission control in a generic heterogeneous network. No specific technology is considered, but it is rather supposed that the two systems have fixed and limited capacity expressed through multiples of some elementary bandwidth units. In UMTS and LTE, due to power control and scheduling respectively, short term capacity variations make such approach hardly feasible. In [11], authors use Semi Markov Decision Process (SMDP) theory to compute optimal user satisfaction for collocated HSDPA (High Speed Downlink Packet Access) – WLAN base stations. In that study coverage areas appear as a main decision issue. Finally, in [12] the authors apply Q-Learning for joint call admission control in GPRS, UMTS and WLAN.

Apart of the fact that we consider different access technologies, with different characteristics, one of the novelties of the proposed approach, is that we also reduce the complexity by building an algorithm that has reduced number of states when compared with other approaches such as [11] and [12]. In general, this property enables faster learning, which is especially important for real-time learning.

On the other hand, the proposed approach can support a distributed implementation at the base stations, assuming that LTE and UMTS technologies are co-located in the same sites, which is also a difference with respect to prior works.

The rest of this paper is organized as follows. In Section II we first give the JRRM-RL (JRL) framework description. Afterwards, in Section III reinforcement learning mechanism for the JRRM model is explained. Simulation parameters and obtained results are included in Section IV. Finally, Section V contains the summary of the main conclusions.

## II. JRRM-RL (JRL) FRAMEWORK

The proposed JRRM mechanism assumes autonomous RL agents assigned to cells. UMTS and LTE are collocated on the same base station in a cell. Notice that these co-located scenarios are very likely to occur in the future since, by reusing UMTS infrastructure, LTE will be more easily deployed at lower expenditures for operators. The agent in each cell works in real-time, independently from the agents in other cells. This agent is responsible for distributing the users between the two technologies (UMTS or LTE). This decision can be taken either at session initiation, or during session lifetime, which could lead to a vertical handover (i.e. a change in the currently assigned RAT), in order to adapt to traffic variations. Fig 1 shows the distributed JRRM model. In each cell mobile users will be directed to LTE or UMTS technology through the RL agent. Based on the current cell conditions (load and interference) and on the previous experience, RL agent will adapt decision policies to increase user satisfaction.

RL is a branch of machine learning where an agent through interaction with environment learns and decides on actions in

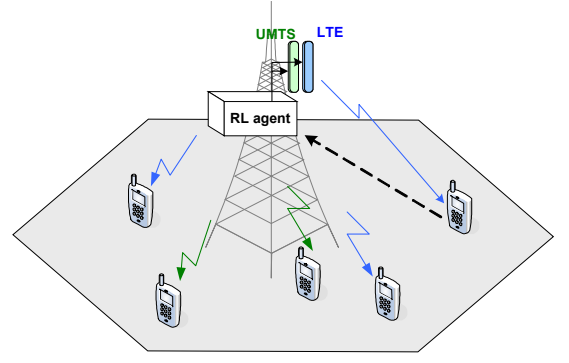


Figure 1. Reinforcement learning for JRRM.

order to maximize some long term reward. The reward is an input that the agent receives from the environment, which represents the quality of the actions taken by the agent, and that should reflect the goals and needs of the system. In this work, we assume that the reward has to reflect the overall user satisfaction with the received service. Specifically, a user is satisfied when it gets a throughput higher than a given threshold  $Th^{lim}$ , no matter which technology it uses. Throughput is evaluated through time and an average is used at instants when reward is calculated. Based on this, we define a limit threshold  $\Theta_{MAX}$  for the maximum tolerable percentage of unsatisfied users in the system.

As aforementioned, in addition to deciding the RAT on the session arrival (incoming requests), the same RL mechanism also decides on session reconfigurations (re-assignment). In particular, after a session has started, the algorithm checks every  $T_{REQ}$  seconds whether the RAT should be changed for a given user or not. Example with three mobiles in Fig 2 illustrates initial and re-assignment decisions by RL agent. Introduction of reconfiguration brings two benefits. First, we avoid long user sessions to stick with one technology in case system conditions change. Second, algorithm learns faster (because the RL mechanism observes and learns based on each request), which is of high importance for such real-time mechanism.

## III. REINFORCEMENT LEARNING FOR JRL

In this paper we use *actor-critic* learning [3], as this method requires minimal computation in order to select actions and can learn optimal probabilities of selecting various actions. The *actor* is the entity that selects actions. The *critic* is a state-value function that takes the form of Temporal Difference (TD) error

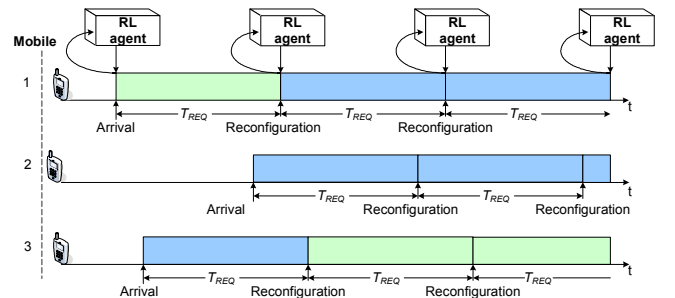


Figure 2. JRL: Session activations and reconfigurations.

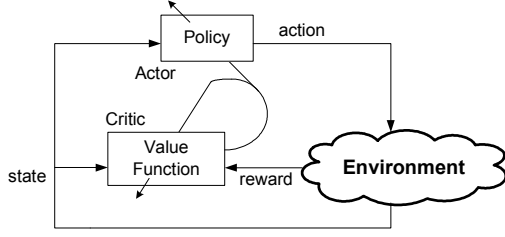


Figure 3. Actor-critic RL: the agent-environment interaction.

and “criticizes” the actions made by the actor: after each action taken it evaluates the new state to determine whether the results have gone better or worse than expected. This is carried out based on the interaction with the environment through a *reward* function. The general procedure is illustrated in Fig 3. In the following, we explain the specific states, actors, learning engine and reward used for the proposed JRL.

#### A. States

Each cell in the system has one RL agent to handle sessions. The agent has nine states implemented (as in Fig. 4-left) that reflect the current conditions (dissatisfaction level  $P_{dis}$ ) in each of the two technologies. These conditions can be: complete satisfaction in UMTS/LTE, dissatisfaction lower than a limit threshold  $\Theta_{MAX}$ , or dissatisfaction higher than the threshold value.

In each state an actor decides on the action selection for a user. Each action corresponds to a RAT assignment (i.e. UMTS or LTE). The initial nine state set (presented in Fig. 4-left) corresponds to users with the same profile – same service and similar performance requirements (QoS requirements: throughput, delay, etc.). In a more general case users can be grouped under different Profile Groups (PGs). In that case the algorithm will contain as many nine state sets as there are different user profile groups in the system (Fig 4-right). Then, the states are defined by the system conditions (satisfaction) and the profile group that current user belongs to.

#### B. Actors

RL agents use softmax action selection, so probability to select action  $a$  (over  $A$  possible actions) in state  $s_{xx}$  of cell  $j$  is:

$$\pi_j(a, s_{xx}) = \frac{e^{p_j(a, s_{xx})/\tau}}{\sum_{b=1}^A e^{p_j(b, s_{xx})/\tau}}. \quad (1)$$

For each action the corresponding learning parameter is  $p_j(a, s_{xx})$ . Parameter  $\tau$  is a positive constant parameter.

#### C. REINFORCE learning engine

The learning process in JRL algorithm is achieved through the update of the parameters  $p_j(a, s_{xx})$ . We use REINFORCE [13] learning techniques to make this update. In particular, for the action  $a_t$  taken in the state  $s_t$  of the cell  $j$  the learning parameter is updated based on the reward as:

$$p_j(a_t, s_t) \leftarrow p_j(a_t, s_t) + \beta \cdot (r_j - \hat{r}_j) \cdot (1 - \pi_j(a_t, s_t)). \quad (2)$$

Here  $\beta$  is a positive step size parameter, whereas  $r_j$  is the reward estimate at cell  $j$ .

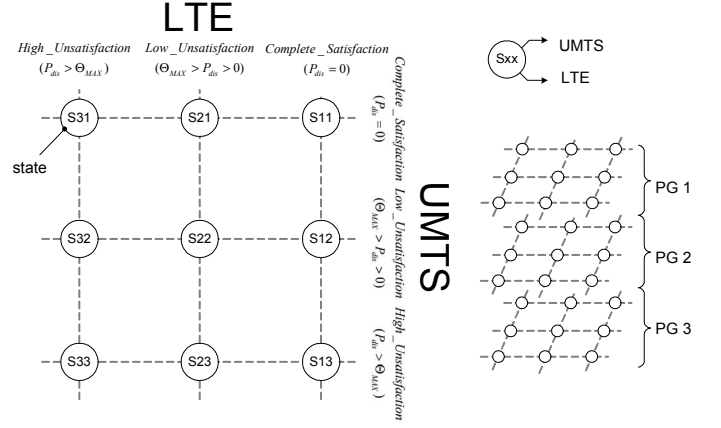


Figure 4. Reinforcement learning model for JRL.

In our JRL mechanism, reward is calculated in each cell periodically (with period  $\Delta RT$ ). The corresponding cumulative reward is then updated after each iteration  $\Delta RT$ :

$$\hat{r}_j \leftarrow \hat{r}_j + \gamma \cdot (r_j - \hat{r}_j) \quad (3)$$

where  $\gamma$  represents the step-size. After this value is obtained the parameters  $p_j(a, s_{xx})$  of the actions taken within that period are updated based on Eq. (2).

#### D. Reward function

The reward function is the evaluation of the actions taken by RL agents through the algorithm’s evolution. The RL algorithm tends to achieve maximization of the reward function  $r$  through time [13]. Consequently, by a proper mapping of the desired QoS goals onto the reward function, these goals can be optimized. In particular, in this work we define the reward for cell  $j$  as:

$$r_j = \frac{n_{active,j}^{satisfied}}{n_{active,j}} - k_{LTE} \cdot \hat{\Theta}_j^{LTE} - k_{UMTS} \cdot \hat{\Theta}_j^{UMTS} \quad (4)$$

The first term of this function reflects the user satisfaction and is computed as the fraction of satisfied users  $n_{active,j}^{satisfied}$  with respect to the overall number of active users  $n_{active,j}$  in the cell  $j$ .

The second and the third terms have been introduced empirically in order to have better performances. We subtracted constant  $k_{LTE}$  or  $k_{UMTS}$  in case dissatisfaction is higher than  $\Theta_{MAX}$  in LTE or UMTS respectively; that is:

$$\hat{\Theta}_j^{LTE} = \begin{cases} 0, & \Theta_j^{LTE} < \Theta_{MAX} \\ 1, & \Theta_j^{LTE} > \Theta_{MAX} \end{cases} \quad \hat{\Theta}_j^{UMTS} = \begin{cases} 0, & \Theta_j^{UMTS} < \Theta_{MAX} \\ 1, & \Theta_j^{UMTS} > \Theta_{MAX} \end{cases} \quad (5)$$

## IV. SIMULATION RESULTS

The simulated environment includes a base station with two 5MHz frequency bands, one used for one carrier of WCDMA/UMTS and the other for OFDMA/LTE. We suppose the reuse factor 3 for the LTE, so that only 8 out of 24 frequency chunks (resource blocks) are assigned to one cell as active. The remaining 16 chunks would be used by neighboring cells. The parameters used in simulations are given in table I.

TABLE I. SIMULATION PARAMETERS

Parameter	Value / Description
1 Cell area	Hexagonal (Radius=0.5 km)
Path loss	$128.1 + 37.6 \cdot \log_{10}(d[\text{km}])$ dB
Shadow fading	Log-normal, std=8dB
Base station shadowing correlation	0.5
User distribution	Uniform
Terminal speed	0
Nominal bandwidth ( $w$ )	5 MHz per RAT
Background noise in downlink ( $P_N$ )	-102 dBm (One RB $P_N/24$ )
Max. base station Power ( $P_{T_i}^{MAX}$ )	43 dBm (per 5MHz)
<b>UMTS</b>	
( $E_b/N_0$ ) target	5 dB
Orthogonality factor	0.4
Pilot + Control	30 dBm
Max. Trans. Power per Mobile ( $P_{T_i}^{MAX}$ )	30 dBm
<b>LTE</b>	
Multipath Fading	3GPP-Case 3 [14]
Frame length	1ms (TTI=0.5ms)
Subcarrier spacing ( $\Delta f$ )	15 kHz
Resource block (RB)	180 kHz (12 x $\Delta f$ )
Subcarrier symbol rate	7 symbols / TTI
L1/L2 Overhead	3 / 14 symbols [15]
Number of RBs	24 (per 5MHz)
Frequency reuse factor	3 (8/24 RBs per cell)
AMC level [16]	QPSK $1/3, 1/2, 2/3$ 16-QAM $1/2, 2/3, 5/6$ 64-QAM $2/3, 5/6$
Scheduler	Proportional fair queuing [17]
<b>Session / System parameters</b>	
System load ( $\lambda/\mu$ )	from 0 to 150
Average session duration	expon. distr., mean 240 sec
Average time between sessions	expon. distr., mean 240 sec
Re-assignment interval ( $T_{REQ}$ )	30 sec
Generated data rate per user ( $R_b$ )	64 kbps (CBR)
Throughput satisfaction threshold ( $Th^{LM}$ )	$0.95 \cdot R_b$
Dissatisfaction threshold ( $\Theta_{MAX}$ )	5 %
<b>RL</b>	
Reward update ( $\Delta RT$ )	400 ms
Parameter $\beta$	2.0
Parameter $\tau$	0.2
Parameter $\gamma$	0.6
$k_{LTE}$	0.2
$k_{UMTS}$	0.4

Evaluation is carried out in the downlink and all users use the same service. Each simulation length was  $10^4$  sec.

In order to assess the performance of the applied RL mechanism, we use some baseline solutions to compare with:

- **Separate UMTS / LTE:** Performance of the two RATs is tested separately in the same scenarios. This is done to get an insight into the average potential of each of the two RATs.
- **Classic load balancing (LB):** This approach assigns an incoming session to the RAT with the lowest load. We use a version of the approach in [18] adapted to the two RATs in this study. In UMTS downlink the load is measured as the transmitted power divided by the maximum power available. In LTE, load is measured as the average number of used chunks over the maximum number of available chunks.
- **Satisfaction-based load balancing (SLB):** We also do tests with load balancing that monitors satisfaction in each RAT. This model assigns sessions to the RAT with the highest satisfaction probability. In case of the same satisfaction conditions a session is assigned to LTE – additional tests, not

presented due to space limits, show that this gives better results when compared with the possibility to randomly assign a RAT or to assign it to UMTS.

In both LB and SLB we introduce re-assignment requests in the same manner we do it for the JRL, so that sessions may be reconfigured after  $T_{REQ}$  seconds.

First results in Fig. 5 show the performance in terms of user satisfaction probability with varying average load. Results demonstrate better performance obtained by means of the JRL algorithm over classical LB and SLB approaches. In particular, assuming a minimum satisfaction probability of 95% ( $1 - \Theta_{MAX}$ ) to determine capacity limits, the maximum allowable average load for JRL is around 111 Erlangs, whereas for LB and SLB it is 91 and 101 Erlangs respectively, thus having a capacity improvement of ~22% and ~10%.

In order to analyze in more detail the reasons for the better performance of JRL algorithm, in the following some results are shown under dynamic conditions in which the average load changes along simulation time. We set four temporal stages (S-I to S-IV) in which average load is 70, 110, 50 and 100 Erlangs respectively. Each stage is 2500 sec long.

Distribution of users between the two RATs is presented in Fig. 6-a. The probabilities to allocate the load between LTE and UMTS are equal for JRL at the beginning. As load increases, RL perceives the capacity limits of UMTS and starts loading LTE with a higher portion of users. JRL shows sporadic satisfaction degradation present with high cell loads (S-II) but to a lower extent than with LB and SLB (see Fig. 6-d). The learning mechanism demonstrates the capability to react and adapt fast to system changes.

LB mechanism demonstrates weakness in both S-II and S-IV due to the incapability to make good decisions for high loads. The reason for this is that the assignment to the less loaded RAT does not always result in more user satisfaction. In fact, even if the load in UMTS and LTE is similar for the LB case (see Fig. 6-b-c), user satisfaction is very different (see Fig. 6-e-f). SLB mechanism gives better results than the LB, but threatens to get into the oscillatory behavior when system load is high. An example for this is between 3500 and 4000 sec in Fig. 6e-f, where a lot of users start to be moved alternatively between the two RATs. This is due to the separate tracking of user satisfaction per RAT that can result in instable behavior for very high load. On the contrary, the proposed JRL mechanism, with the ability to track the overall satisfaction probability resulting from the different actions, is able to keep the satisfaction probability at good values in all the cases.

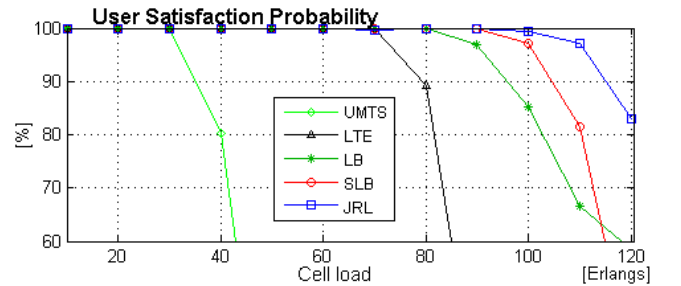


Figure 5. User satisfaction probability.



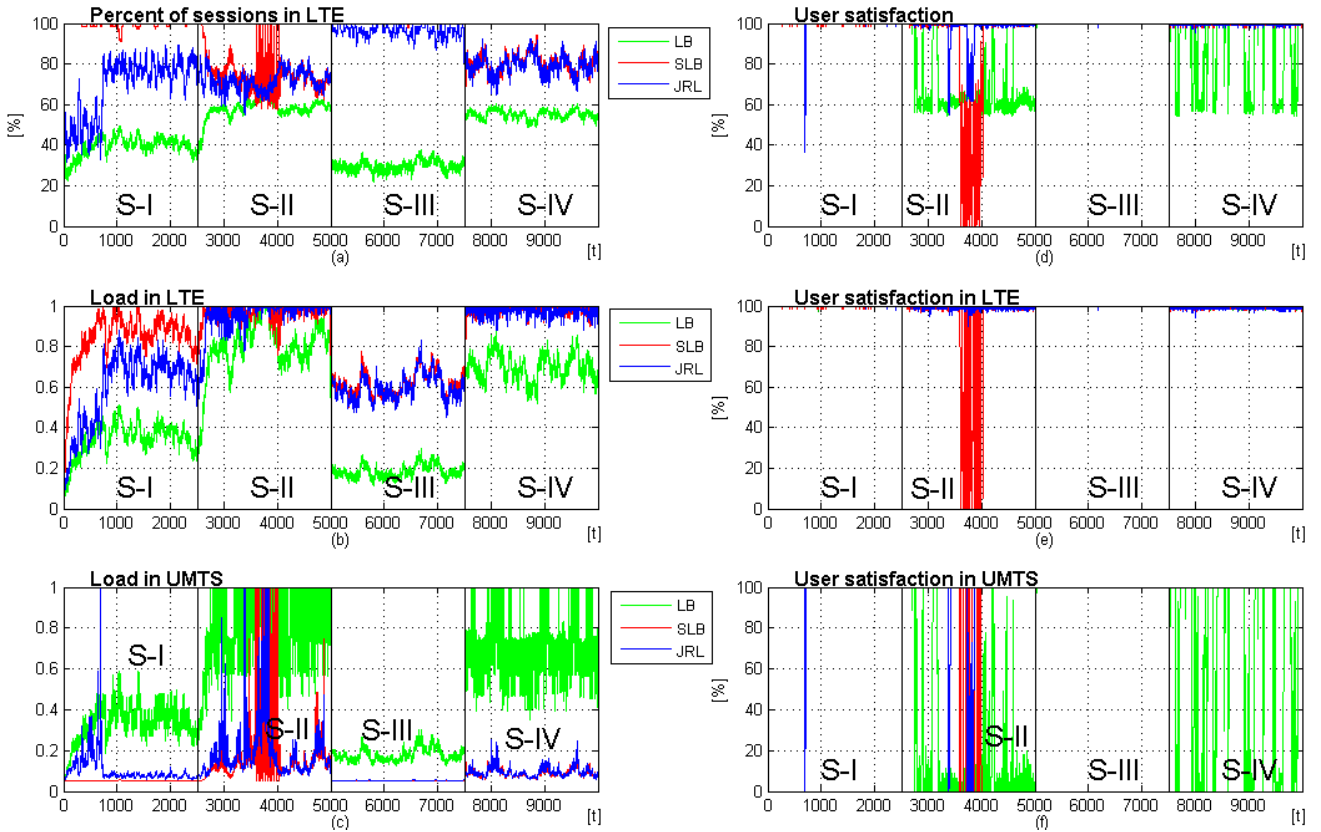


Figure 6. System adaptation: percentage of sessions in LTE in a), load in b) LTE and c) UMTS, user satisfaction probability d) overall, e) in LTE, f) in UMTS.

## V. CONCLUSION

This paper has proposed a dynamic JRRM mechanism based on reinforcement learning for LTE-UMTS coexistence scenarios. Simulation results have shown that the proposed mechanism outperforms both classic load balancing solution and load balancing based on user satisfaction probability. This kind of adaptive JRRM algorithms may serve as a tool to improve the spectral efficiency in mobile communications for the future technology transition period while ensuring user satisfaction. In the future work we plan to test the algorithm with more diverse services, spectrum conditions and with multiple cell scenarios.

## REFERENCES

- [1] A. Tölli, P. Hakanin, H. Holma, "Performance evaluation of common radio resource management (CRRM)", IEEE ICC 2002, pp. 3429-3433
- [2] E. Dahlman, S. Parkvall, J. Sköld, P. Beming, *3G evolution: HSPA and LTE for mobile broadband*, Elsevier, 2007
- [3] R. S. Sutton, A.G. Barto, *Reinforcement Learning: An Introduction*, A Bradford Book, MIT Press, Cambridge, MA 1998
- [4] R.W. Thomas, D.H. Friend, L.A. DaSilva, A.B. MacKenzie, "Cognitive networks: adaptation and learning to achieve end-to-end performance objectives", *IEEE Comm. Magazine*, vol 44, issue 12, Dec. 2006, pp.51-57
- [5] N. Vucevic, J. Perez-Romero, O. Sallent, R. Agustí, "Reinforcement learning for dynamic spectrum management in WCDMA", 16th Telecommunications Forum (TELFOR 2008), pp. 183-186
- [6] O.E. Falowo, H.A. Chan, "Joint call admission control algorithms: requirements, approaches, and design considerations", *Computer Communications*, vol. 31, issue 6, 18 April 2008, pp. 1200-1217
- [7] J. Pérez-Romero, O. Sallent, R. Agustí, "On evaluating beyond 3G radio access networks: architectures, approaches and tools", IEEE VTC Spring 2005, pp. 2964-2968
- [8] J. Perez-Romero, et al, "Network-controlled cell-breathing for capacity improvement in heterogeneous CDMA/TDMA scenarios", IEEE WCNC 2006, pp. 36-41
- [9] L. Giupponi, R. Agustí, J. Perez-Romero, O. Sallent Roig, "A Novel Approach for Joint Radio Resource Management Based on Fuzzy Neural Methodology", *IEEE Tran. on Vehicular Technology*, vol 57, issue 3, May 2008, pp. 1789 - 1805
- [10] E. Alexandri, G. Martinez, D. Zeghlache, "Adaptive joint call admission control and access network selection for multimedia wireless systems", WPMC 2002, pp.1390 - 1394
- [11] M. Coupechoux, J.-M. Kelif, P. Godlewski, "Network Controlled Joint Radio Resource Management for Heterogeneous Networks", IEEE VTC Spring 2008, pp. 1771 - 1775
- [12] Z. Feng, Y. Xue, Y. Zhang, X. Tao, "Autonomic Joint Session Admission Control using Reinforcement Learning", ICT-MobileSummit 2008
- [13] R.J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning", *Machine Learning*, vol. 8, 1992
- [14] 3GPP TS 25.104, *Base Station (BS) radio transmission and reception (FDD)*, Release 8 (v.8.4.1), Sept. 2008, p. 75 (Annex B.1)
- [15] M.E. Ortega Lerida, *Adaptive radio resource management for voip and data traffic in 3GPP LTE networks*, Master thesis, Royal institute of technology (KTH), Stockholm, Sweden, March 2008
- [16] R. Schoenen, W. Zirwas, B.H. Walke, "Capacity and coverage analysis of a 3GPP-LTE multihop deployment scenario", Communications Workshops, ICC Workshops 2008, pp. 31-36
- [17] C. Wengerter, J. Ohlhorst, A.G.E. von Elbwart, "Fairness and throughput analysis for generalized proportional fair frequency scheduling in OFDMA", IEEE VTC Spring 2005, pp. 1903-1907
- [18] X. Gelabert, J. Pérez-Romero, O. Sallent, R. Agustí, "On the Suitability of Load Balancing Principles in Heterogeneous Wireless Access Networks", WPMC 2005, pp. 1503-1507